Advanced Information Systems Technology

OceanXtremes: Oceanographic Data-Intensive Anomaly Detection and Analysis Portal

Completed Technology Project (2015 - 2017)

Project Introduction

Anomaly detection is a process of identifying items, events or observations, which do not conform to an expected pattern in a dataset or time series. Current and future missions and our research communities challenge us to rapidly identify features and anomalies in complex and voluminous observations to further science and improve decision support. Given this data intensive reality, we propose to develop an anomaly detection system, called OceanXtremes, powered by an intelligent, elastic Cloud-based analytic service backend that enables execution of domain-specific, multi-scale anomaly and feature detection algorithms across the entire archive of ocean science datasets. A parallel analytics engine will be developed as the key computational and data-mining core of OceanXtreams' backend processing. This analytic engine will demonstrate three new technology ideas to provide rapid turn around on climatology computation and anomaly detection: 1. An adaption of the Hadoop/MapReduce framework for parallel data mining of science datasets, typically large 3 or 4 dimensional arrays packaged in NetCDF and HDF. 2. An algorithm profiling service to efficiently and cost-effectively scale up hybrid Cloud computing resources based on the needs of scheduled jobs (CPU, memory, network, and bursting from a private Cloud computing cluster to public cloud provider like Amazon Cloud services). 3. An extension to industry-standard search solutions (OpenSearch and Faceted search) to provide support for shared discovery and exploration of ocean phenomena and anomalies, along with unexpected correlations between key measured variables. We will use a hybrid Cloud compute cluster (private Eucalyptus on-premise at JPL with bursting to Amazon Web Services) as the operational backend. The key idea is that the parallel data-mining operations will be run 'near' the ocean data archives (a local 'network' hop) so that we can efficiently access the thousands of (say, daily) files making up a three decade time-series, and then cache key variables and pre-computed climatologies in a high-performance parallel database. OceanXtremes will be equipped with both web portal and web service interfaces for users and applications/systems to register and retrieve oceanographic anomalies data. By leveraging technology such as Datacasting (Bingham, et.al, 2007), users can also subscribe to anomaly or 'event' types of their interest and have newly computed anomaly metrics and other information delivered to them by metadata feeds packaged in standard Rich Site Summary (RSS) format. Upon receiving new feed entries, users can examine the metrics and download relevant variables, by simply clicking on a link, to begin further analyzing the event. The OceanXtremes web portal will allow users to define their own anomaly or feature types where continuous backend processing will be scheduled to populate the new user-defined anomaly type by executing the chosen data mining algorithm (i.e. differences from climatology or gradients above a specified threshold). Metadata on the identified anomalies will be cataloged including temporal and geospatial profiles, key physical metrics, related observational artifacts and other relevant metadata to facilitate discovery, extraction, and visualization. Products created by the anomaly detection

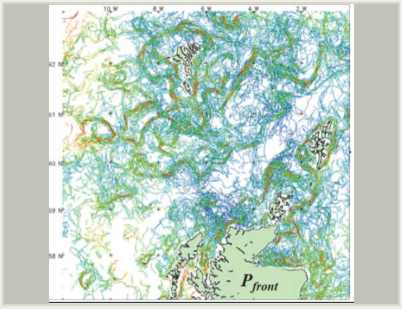


Illustration of future OceanXtremes analysis capability showing sea surface temperature frontal probability for AVHRR imagery (warmer colors indicated higher frontal probability and persistence).

Table of Contents

OceanXtremes: Oceanographic Data-Intensive Anomaly Detection and Analysis Portal

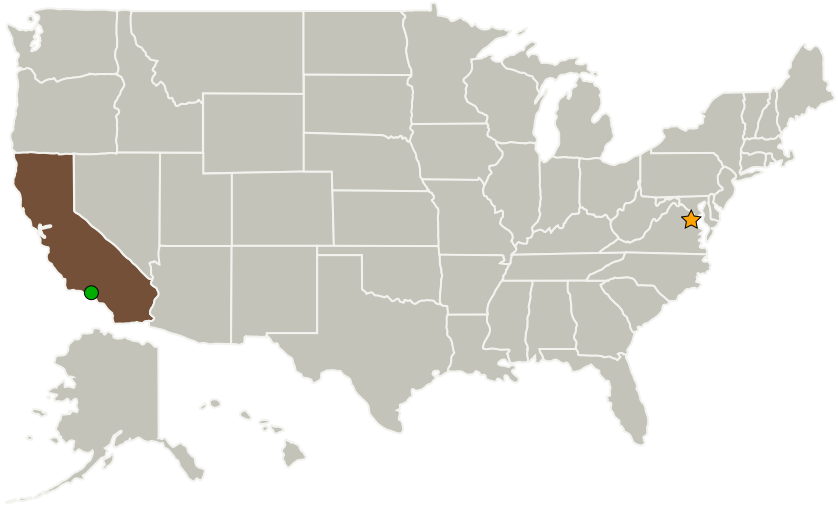
Completed Technology Project (2015 - 2017)

algorithm will be made explorable and subsettable using Webification (Huang, et.al, 2014) and OPeNDAP (http://opendap.org) technologies. Using this platform scientists can efficiently search for anomalies or ocean phenomena, compute data metrics for events or over time-series of ocean variables, and efficiently find and access all of the data relevant to their study (and then download only that data).

Anticipated Benefits

SWOT

Primary U.S. Work Locations and Key Partners



Organizations Performing Work	Role	Type	Location
★NASA Headquarters(HQ)	Lead Organization	NASA Center	Washington, District of Columbia
●Jet Propulsion Laboratory(JPL)	Supporting Organization	NASA Center	Pasadena, California

Primary U.S. Work Locations
California

Organizational Responsibility

Responsible Mission Directorate:

Science Mission Directorate (SMD)

Lead Center / Facility:

NASA Headquarters (HQ)

Responsible Program:

Advanced Information Systems Technology

Project Management

Program Director:

Pamela S Millar

Program Manager:

Jacqueline J Le Moigne

Principal Investigator:

Thomas Huang

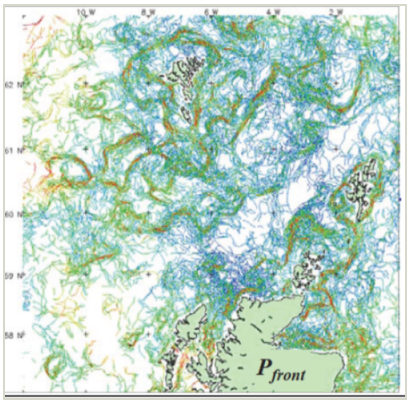
Co-Investigators:

Toshio M Chin
Edward M Armstrong
Brian D Wilson
Victor Zlotnicki
Karen R Piggee
George W Chang
Tong Lee

OceanXtremes: Oceanographic Data-Intensive Anomaly Detection and Analysis Portal

Completed Technology Project (2015 - 2017)

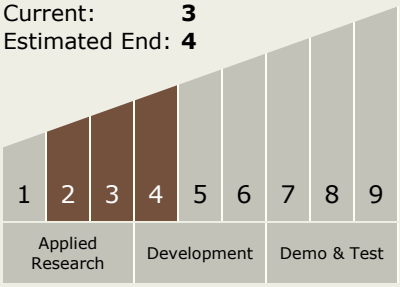
Images



Untitled
Illustration of future OceanXtremes analysis capability showing sea surface temperature frontal probability for AVHRR imagery (warmer colors indicated higher frontal probability and persistence). *(https://techport.nasa.gov/image/16562)*

Technology Maturity (TRL)

Start: **2**
Current: **3**
Estimated End: **4**



Technology Areas

Primary:

- TX11 Software, Modeling, Simulation, and Information Processing
  - TX11.4 Information Processing
    - TX11.4.2 Intelligent Data Understanding

Target Destination

Earth